POTENTIAL OF DATA MINIMIZATION TECHNIQUES APPROACH IN IMPROVING DNA SEQUENCES ALIGNMENT AND COMPARISON

RESEARCH MANAGEMENT INSTITUTE (RMI)
UNIVERSITI TEKNOLOGI MARA
40450 SHAH ALAM, SELANGOR
MALAYSIA

BY :

SYED ABDUL MUTALIB AL JUNID
MUHAMMAD ADIB BIN HARON
ZULKIFLI BIN ABD MAJID
ZULKIFLI BIN OTHMAN

DECEMBER 2012

Contents

2. Letter of Offer (Research Grant)

UNIVERSITI TEKNOLOGI MARA

PEMENANG Anugerah Kualiti Perdana Menteri 2008

Surat Kami : 600-RMI/ST/DANA 5/3/Dst (314 /2011)
Tarikh : 27 Mei 2011

En Syed Abdul Mutalib Al Junid
Fakulti Kejuruteraan Elektrik
Universiti Teknologi MARA
40450 Shah Alam

Y. Brs. Profesor /Tuan/Puan

KELULUSAN PERMOHONAN DANA KECEMERLANGAN 05/2011

Tajuk Projek	Potential Of Data Minimization Technique Approach In Improving DNA Sequences Alignment And Comparison
Kod Projek	600-RMI/ST/DANA 5/3/Dst (314 /2011)
Kategori Projek	Kategori F (2011)
Tempoh	01 Jun 2011 – 31 Mei 2012 (12 bulan)
Jumlah Peruntukan	RM 8,000.00
Ketua Projek	En Syed Abdul Mutalib Al Junid

Dengan hormatnya perkara di atas adalah dirujuk.

2. Sukacita dimaklumkan pihak Universiti telah meluluskan cadangan penyelidikan Y. Brs. Profesor/tuan/puan untuk membiayai projek penyelidikan di bawah Dana Kecemerlangan UiTM.

3. Bagi pihak Universiti kami mengucapkan tahniah kepada Y. Brs. Profesor/tuan/puan kerana kejayaan ini dan seterusnya diharapkan berjaya menyiapkan projek ini dengan cemerlang.

4. Peruntukan kewangan akan disalurkan melalui tiga (3) peringkat berdasarkan kepada laporan kemajuan serta kewangan yang mencapai perbelanjaan lebih kurang 50% dari peruntukan yang diterima.

Peringkat Pertama	20%
Peringkat Kedua	40%
Peringkat Ketiga	40%

5. Untuk tujuan mengemaskini, pihak Y. Brs. Profesor/tuan/puan adalah diminta untuk melengkapkan semula kertas cadangan penyelidikan sekiranya perlu, mengisi borang setuju terima projek penyelidikan dan menyusun perancangan semula bajet yang baru seperti yang diluluskan. Sila lihat lampiran bagi tatacara tambahan untuk pengurusan projek.

Sekian, harap maklum.

"SELAMAT MENJALANKAN PENYELIDIKAN DENGAN JAYANYA"

Yang benar

MUSTAFAR KAMAL HAMZAH
Ketua Penyelidikan (Sains dan Teknologi)

/RS

Penolong Naib Canselor (Penyelidikan) [illegible] Bahagian Penerbitan [illegible] Penolong Pentadbiran [illegible]
Bahagian Penyelidikan [illegible] Bahagian Sokongan ICT [illegible] Fax 603-5544 2099 / 2767
Bahagian Perundingan [illegible] Bahagian Sains [illegible] Unit Kewangan Zon 17 603-5544 3404
Bahagian Inovasi [illegible] Pejabat Am [illegible] 603-5523 1386

Research Management Institute (RMI) Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
http://www.rmi.uitm.edu.my

5. Report

5.1 Proposed Executive Summary

Nowadays, the demand for advanced and high performance computational method for comparing and searching biological sequences have increased according to exponential growth rate of biological sequences database. Besides this demand, the requirement for high performance and sensitive comparison and alignment tools have also increased after the advantage of the system for defining the solution relates to Deoxyribonucleic Acid (DNA), human genomes and molecular biology figured out through bioinformatics study. Thereby, this research is an attempt to study and investigate the potential of data minimization technique approach in improving the process of aligning and comparing DNA sequences for determine the region of similarity between two or more DNA sequences. Specifically, the research objectives are i) to study various DNA sequences alignment algorithms characteristics, architectures and techniques such as scoring functions, optimal alignment and gap; and ii) to investigate the potential of data minimization approach in minimizing the size of data towards discovering the region of similarity in DNA sequences alignment and comparison and iii) to analyze the optimal sequences alignment properties for molecular biology and bioinformatics application. This study will initially involve algorithm modification, mathematical formulation, logic design and simulation for data minimization and existing DNA sequences alignment algorithm. At the end of the study, it is expecting that we will discover the potential of data minimization technique in minimizing the size of data towards aligning DNA sequences using basic dynamic programming concept.

5.2 Enhanced Executive Summary

Improved data minimization technique to optimize the length of DNA sequence and alignment result characters representation presents in this report. The primary objective is to improve and optimize data representation for DNA sequences alignment and result character. The proposed design change in algorithm and architecture is highlights in this report. Algorithm design based on binary equivalent method is using to obtain the optimal size of characters representation. The code had written, compiled and simulated using Altera Quartus II Version 9.0 EDA tools. While, Verilog Hardware Description Language (HDL) and Altera Cyclone II EP2C35 FPGA used as coding language and target device respectively. In addition, the structural modelling technique for reduce the design complexity is used. Simulation result showed that the improved data minimization technique takes 50% more memory compared to existing data minimization, but it covers 6 DNA sequences and alignment result characters.